

# Russia Enterprise Surveys Data Set

## 1. Introduction

1. This document provides additional information on the data collected in Russia during calendar years 2008/2009 as part of the fourth round of the Business Environment and Enterprise Performance Survey (BEEPS IV), a joint initiative of the World Bank Group (“WB”) and the European Bank for Reconstruction and Development (“EBRD”). It is an enterprise survey whose objective is to gain an understanding of firms’ perception of the environment in which they operate. The survey was until now administered three times at three years interval. This has added an important element of dynamics in the study of business environment in transition countries.

The 2008 survey was restructured to improve cross-country comparability and to make it compatible with the Enterprise Surveys the Enterprise Analysis Unit of the World Bank has been implementing in the past two years in other regions of the world.

The objective of the survey is to obtain feedback from enterprises in client countries on the state of the private sector as well as to help in building a panel of enterprise data that will make it possible to track changes in the business environment over time, thus allowing, for example, impact assessments of reforms.

Through interviews with firms in the manufacturing and services sectors, the survey will assess the constraints to private sector growth and create statistically significant business environment indicators that are comparable across countries.

The report outlines and describes the sampling design of the data, the data set structure as well as additional information that may be useful when using the data, such as information on non-response cases and the appropriate use of the weights.

## 2. Sampling Structure

2. The sample for Russia was selected using stratified random sampling, following the methodology explained in the Sampling Manual<sup>1</sup>. Stratified random sampling<sup>2</sup> was preferred over simple random sampling for several reasons<sup>3</sup>:

a. To obtain unbiased estimates for different subdivisions of the population with some known level of precision.

b. To obtain unbiased estimates for the whole population. The whole population, or universe of the study, is the non-agricultural economy. It comprises: all manufacturing sectors according to the group classification of ISIC Revision 3.1: (group D), construction sector (group F), services sector (groups G and H), and transport, storage, and communications sector (group I). Note that this definition excludes the following

---

<sup>1</sup> The complete text can be found at [http://www.enterprisesurveys.org/documents/Implementation\\_note.pdf](http://www.enterprisesurveys.org/documents/Implementation_note.pdf)

<sup>2</sup> A stratified random sample is one obtained by separating the population elements into non-overlapping groups, called strata, and then selecting a simple random sample from each stratum. (Richard L. Scheaffer; Mendenhall, W.; Lyman, R., “Elementary Survey Sampling”, Fifth Edition).

<sup>3</sup> Cochran, W., 1977, pp. 89; Lohr, Sharon, 1999, pp. 95

sectors: financial intermediation (group J), real estate and renting activities (group K, except sub-sector 72, IT, which was added to the population under study), and all public or utilities-sectors.

c. To make sure that the final total sample includes establishments from all different sectors and that it is not concentrated in one or two of industries/sizes/regions.

d. To exploit the benefits of stratified sampling where population estimates, in most cases, will be more precise than using a simple random sampling method (i.e., lower standard errors, other things being equal.)

e. Stratification may produce a smaller bound on the error of estimation than would be produced by a simple random sample of the same size. This result is particularly true if measurements within strata are homogeneous.

f. The cost per observation in the survey may be reduced by stratification of the population elements into convenient groupings.

3. Three levels of stratification were used in this country: industry, establishment size, and oblast (region). The original sample design with specific information of the industries and regions chosen is described in Appendix E.

4. Industry stratification was designed in the way that follows: the universe was stratified into

- Manufacturing sectors “15”, “18”, “24”, “27-28”, “29”, and “Other Manufacturing”
- Services sectors “52” and “Other Services”

*Sector 15* had a target of 179 interviews, *Sector 18* had a target of 164 interviews, *Sector 24* had a target of 162 interviews, *Sector 27-28* had a target of 164 interviews, *Sector 29* had a target of 167 interviews, *Sector 52* had a target of 144 interviews, *Other Manufacturing* had a target of 141 interviews and *Other Services* had a target of 139 interviews.

5. Size stratification was defined following the standardized definition for the rollout: small (5 to 19 employees), medium (20 to 99 employees), and large (more than 99 employees)<sup>4</sup>. For stratification purposes, the number of employees was defined on the basis of reported permanent full-time workers. This seems to be an appropriate definition of the labor force since seasonal/casual/part-time employment is not a common practice, except in the sectors of construction and agriculture.

6. Regional stratification was defined in seven regions. These regions are North West, Central, South, Ural, Siberia, Volgo-Viatsky, and Far East.

### **3. Sampling implementation**

7. Given the stratified design, sample frames containing a complete and updated list of establishments for the selected regions were required. Great efforts were made to obtain the best source for these listings. However, the quality of the sample frames was

---

<sup>4</sup> The panel firms from BEEPS 2005 with less than 5 employees are included in the 5 to 19 strata.

not optimal and, therefore, some adjustments were needed to correct for the presence of ineligible units. These adjustments are reflected in the weights computation (*see below*).

8. For most countries covered in BEEPS IV two sample frames were used. The first frame for Russia was compiled in 2007 and obtained from the National Statistics Agency [GosKomStat]. The frame did not show the number of employees for establishments. It did however, show turnover values. Estimates were obtained which related turnover to size. They suggested that to qualify for the usual size criterion the turnover should be at least 1 million roubles. In addition, as Russia is a very large country spanning eleven time zones, the frame would cover many cities. Therefore, for cost efficiency reasons, it was decided that an extract should be purchased that covered only an agreed set of cities for establishments with turnover in excess of 1 million roubles. That extract, selected to instructions of the TNS statistical team, was sent to the TNS statistical team in London to select the establishments for interview. The second frame, supplied by the World Bank/EBRD, consisted of enterprises interviewed in BEEPS 2005. The clients required that the attempts should be made to re-interview establishments responding to the BEEPS 2005 survey where they were within the selected geographical regions and met eligibility criteria. That sample is referred to as the Panel.

9. The quality of the frame was assessed at the onset of the project. The frame proved to be useful though it showed positive rates of non-eligibility, repetition, non-existent units, etc. These problems are typical of establishment surveys, but given the impact these inaccuracies may have on the results, adjustments were needed when computing the appropriate weights for individual observations. The percentage of confirmed non-eligible units as a proportion of the total number of contacts to complete the survey was 44% (2468 out of 5559 establishments).

## Sample Frame Russia

### Fresh

Source: GosKomStat - National Statistics Agency, 2007

Region	Size	Sector							Other Manufacturing	Other Services	Grand Total
		15	18	24	27-28	29	52				
North West	<60m RUB	166	74	102	377	367	1,827	1,353	10,548	14,815	
	60m-299m RUB	77	15	42	79	84	137	397	1,821	2,652	
	300m RUB +	62	4	17	21	33	101	280	1,570	2,088	
North West Total		305	93	161	477	484	2,065	2,030	13,939	19,554	
Central	<60m RUB	591	330	448	760	1,086	6,087	4,473	49,309	63,083	
	60m-299m RUB	166	33	112	134	199	722	1,567	11,565	14,498	
	300m RUB +	155	16	56	52	57	226	1,262	5,210	7,035	
Central Total		912	379	616	946	1,342	7,035	7,302	66,084	84,616	
South	<60m RUB	103	40	35	132	124	1,611	448	4,584	7,076	
	60m-299m RUB	49	1	6	19	31	186	146	1,550	1,989	
	300m RUB +	22	2	3	8	11	34	70	452	602	
South Total		174	43	44	160	166	1,831	663	6,586	9,667	
Ural	<60m RUB	70	37	50	171	202	893	498	5,409	7,330	
	60m-299m RUB	28	4	4	35	33	71	143	1,064	1,383	
	300m RUB +	20	1	7	32	34	65	110	532	800	
Ural Total		119	42	62	237	268	1,029	751	7,005	9,513	
Siberia	<60m RUB	119	35	51	128	143	1,439	561	4,935	7,412	
	60m-299m RUB	30	0	11	23	29	206	155	873	1,327	
	300m RUB +	23	1	5	18	19	9	50	645	772	
Siberia Total		172	36	67	170	191	1,654	767	6,454	9,511	
Volgo-Viatsky	<60m RUB	158	86	142	303	385	2,529	1,064	10,230	14,897	
	60m-299m RUB	75	6	41	67	80	200	409	2,818	3,695	
	300m RUB +	47	0	39	19	43	92	255	1,604	2,099	
Volgo-Viatsky Total		280	92	221	389	508	2,821	1,728	14,652	20,691	
Far East	<60m RUB	46	15	6	20	20	705	184	1,983	2,979	
	60m-299m RUB	11	0	1	4	3	49	26	350	444	
	300m RUB +	17	0	0	1	1	16	2	292	329	
Far East Total		74	15	7	25	24	770	212	2,625	3,752	
Grand Total		2,036	700	1,178	2,404	2,983	17,205	13,453	117,345	157,304	

### Panel

Source: BEEPS 2005

Region	Employees	Sector							Other Manuf.	Other Serv.	G. Total
		15	18	24	27-28	29	52				
North West	< 5								2	2	
	5 - 19					1	5	5	11	22	
	20 - 99			1		1	2	1	11	16	
	100+	2		1				2	6	11	
North West Total		2		2		2	7	8	30	51	
Central	< 5		2				11	1	4	18	
	5 - 19	1	5	1		1	22	12	24	66	
	20 - 99	8					8	6	43	65	
	100+	3	3	1		3	6	17	28	61	
Central Total		12	10	2		4	47	36	99	210	
South	< 5							3		3	
	5 - 19						2	1	6	9	
	20 - 99		1	1		1	1	2	8	14	
	100+					1		3	3	7	
South Total			1	1		2	3	9	17	33	
Ural	< 5										
	5 - 19		2				2	1	6	11	
	20 - 99							3	3	6	
	100+	1							3	4	
Ural Total		1	2				2	4	12	21	
Siberia	< 5						4	1		5	
	5 - 19						3	1	2	6	
	20 - 99	2					2	5	7	16	
	100+							2	3	5	
Siberia Total		2					9	9	12	32	
Volgo-Viatsky	< 5		1				1		1	3	
	5 - 19		1				6	3	7	17	
	20 - 99	1	1				1	2	10	15	
	100+					1	1	4	5	11	
Volgo-Viatsky Total		1	3			1	9	9	23	46	
Grand Total		18	16	5		9	77	75	193	393	

### Sectors included in the Sample:

Original Sectors	Manufactures: 15, 18, 24, 27, 28, 29 Services: 52 Residual: 16, 17, 19, 20, 21, 22, 23, 25, 26, 30, 31, 32, 33, 34, 35, 36, 37, 45, 50, 51, 55, 60, 61, 62, 63, 64 and 72.
Added Sectors	No

#### 4. Data Base Structure:

10. The structure of the data base reflects the fact that 3 different versions of the questionnaire were used. The basic questionnaire, the Core Module, includes all common questions asked to all establishments from all sectors (manufacturing, services and IT). The second expanded variation, the Manufacturing Questionnaire, is built upon the Core Module and adds some specific questions relevant to the sector. The third expanded variation, the Services Questionnaire, is also built upon the Core Module and adds to the core specific questions relevant to either retail or IT. Each variation of the questionnaire is identified by the index variable, *a0*.

11. All variables are named using, first, the letter of each section and, second, the number of the variable within the section, i.e. *a1* denotes section *A*, question *1*. Variable names preceded by a prefix “*ECA*” indicate questions used in the previous rollout (2005) and, therefore, they may not be found in the implementation of the rollout in other Countries. All other suffixed variables are global and are present in all country surveys over the world. All variables are numeric with the exception of those variables with an “*x*” at the end of their names. The suffix “*x*” denotes that the variable is alpha-numeric.

12. There are 2 establishment identifiers, *idstd* and *id*. The first is a global unique identifier. The second is a country unique identifier. The variables *a2* (sampling region), *a6a* (sampling establishment’s size), and *a4a* (sampling sector) contain the establishment’s classification into the strata chosen for each country using information from the sample frame. The strata were defined according to the guidelines described above.

13. As noted above, there are 3 levels of stratification: industry, size and region. Different combinations of these variables generate the strata cells for each industry/region/size combination. A distinction should be made between the variable *a4a* and *d1a2* (*industry expressed as ISIC rev. 3.1 code*). The former gives the establishment’s classification into one of the chosen industry-strata, whereas the latter gives the actual establishment’s industry classification (*four digit code*) in the sample frame.

14. All of the following variables contain information from the sampling frame and were defined with the sampling design. They may not coincide with the reality of individual establishments as sample frames may contain inaccurate information. The variables containing the sample frame information are included in the data set for researchers who may want to further investigate statistical features of the survey and the effect of the survey design on their results.

-*a2* is the variable describing sampling regions (oblasts)

-*a6a*: coded using the same standard for small, medium, and large establishments as defined above.

-*a4a*: coded using ISIC codes for the chosen industries for stratification. These codes include most manufacturing industries (15 to 36), and retail, and IT for services (52, and 72 respectively).

-*id2005*: The variable contains the firm ids of the panel firms

-*id2007*: The variable contains the firm ids of the panel firms interviewed in 2007. (available only in Bulgaria, Albania, and Croatia)

15. The surveys were implemented following a 2 stage procedure. In the first stage a screener questionnaire was applied over the phone to determine eligibility and to make appointments; in the second stage, a face-to-face interview took place with the Manager/Owner/Director of each establishment. The variables *a4b* and *a6b* contain the industry and size of the establishment from the screener questionnaire. Variables *a8* to *a11* contain additional information and were also collected in the screening phase.

16. Note that there are additional variables for location (*a3x*), industry (*d1a2*), and size (*l1*, *l6* and *l8*) that reflect more accurately the reality of each establishment. Advance users are advised to use these variables for analytical purposes.

17. Variable *a3x* indicates the actual location of the establishment. There may be divergences between the location in the sampling frame and the actual location, as establishments may be listed in one place but the actual physical location is in another place.

18. Variable *d1a2* indicates the actual ISIC code of the main output of the establishment as answered by the interviewee. This is probably the most accurate variable to classify establishments by activity.

19. Variables *l1*, *l6* and *l8* were designed to obtain a more accurate measure of employment accounting for permanent and temporary employment. Special efforts were made to make sure that this information was not missing for most establishments.

20. Variables *a17x* gives interviewer comments, including problems that occurred during an interview and extraordinary circumstances which could influence results.

## **5. Universe Estimates**

21. Universe estimates for the number of establishments in each cell in Russia were produced for the strict, weak and median eligibility definitions. The estimates were the multiple of the relative eligible proportions.

22. Appendix C shows the overall estimates of the numbers of establishments based on the strict, weak and median relative estimates.

## **6. Weights**

23. Since the sampling design was stratified and employed differential sampling individual observations should be properly weighted when making inferences about the population. Under stratified random sampling unweighted estimates are biased unless sample sizes are proportional to the size of each stratum. With stratification the probability of selection of each unit is, in general, not the same. Consequently, individual observations must be weighted by the inverse of their probability of selection (probability weights or *pw* in Stata.)<sup>5</sup>

24. Special care was given to the correct computation of the weights. Considering the varying quality of the sample frames, it was imperative to accurately adjust the totals within each region/industry/size stratum to account for the presence of ineligible units (the firm discontinued businesses or was unattainable, education or government establishments, establishments with less than 5 employees, no reply after having called in different days of the week and in different business hours, out of order, no tone in the phone line, answering machine, fax line, wrong address or moved away and could not get the new references) The information required for the adjustment was collected in the first stage of the implementation: the screening process. Using this information, each stratum cell of the universe was scaled down by the observed proportion of ineligible units within the cell. Once an accurate estimate of the universe cell (projections) was available, weights were computed using the number of completed interviews. Please, note that panel firms with less than 5 employees were also included in the eligible sample and special coded zero was used in a6a and a6b (sample and screener size) to reflect those cases.

25. For some units it was impossible to determine eligibility because the contact was not successfully completed. Consequently, different assumptions as to their eligibility result in different universe cells' adjustments and in different sampling weights. Three sets of assumptions were considered:

a- Strict assumption: eligible establishments are only those for which it was possible to directly determine eligibility. The resulting weights are included in the variable *w\_strict*.

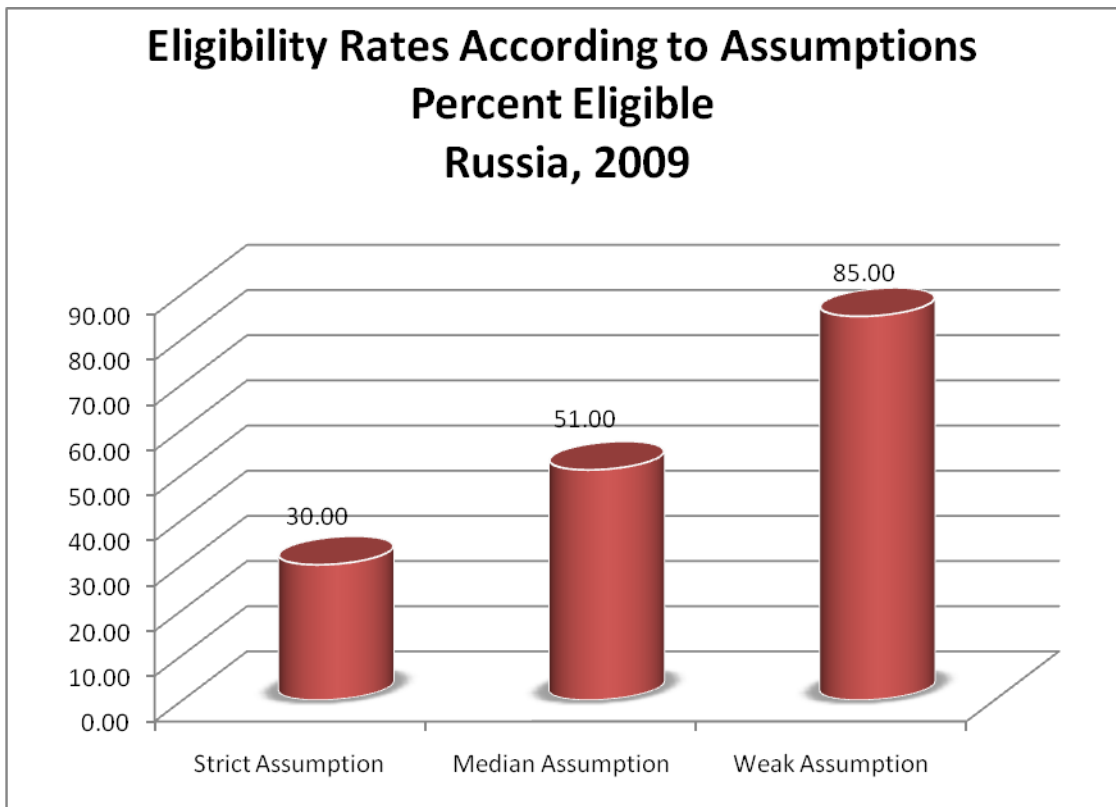
b- Median assumption: eligible establishments are those for which it was possible to directly determine eligibility and those that rejected the screener questionnaire or an answering machine or fax was the only response. The resulting weights are included in the variable *w\_median*.

c- Weak assumption: in addition to the establishments included in points a and b, all establishments for which it was not possible to finalize a contact are assumed eligible. This includes establishments with dead or out of service phone lines, establishments that never answered the phone, and establishments with incorrect addresses for which it was impossible to find a new address. The resulting weights are included in the variable *w\_weak*. Note that under the weak assumption only observed non-eligible units are excluded from universe projections.

The following graph exhibits the different eligibility rates under each set of assumptions.

---

<sup>5</sup> This is equivalent to the weighted average of the estimates for each stratum, with weights equal to the population shares of each stratum.



26. Within each of these assumptions regarding eligibility a pair of weight sets was calculated. The first set of estimates calculated proportions using the raw sample count for each cell. However, the achieved sample numbers in many cells were small. Hence, those eligibility rates, and the adjusted universe cells projections, are subject to relatively large sampling variations. Therefore a second set of more robust estimates (collapsed weights) was also produced. These estimates made use of the multiples of the relative eligibility rates for each industry, size, and region. Those relative rates were based on much larger samples than the individual cells and thus produced values with smaller sampling variations. The data sets include only these robust weights.

Please note that for the purpose of the weights computations all panel firms were considered to be part of the current universe, although technically they are not randomly selected.

### 7. Appropriate use of the weights

27. As discussed above, under stratified random sampling weights should be used when making inferences about the population. Any estimate or indicator that aims at describing some feature of the population should take into account that individual observations may not represent equal shares of the population.



28. However, there is some discussion as to the use of weights in regressions (see Deaton, 1997, pp.67; Lohr, 1999, chapter 11, Cochran, 1953, pp.150). There is not strong large sample econometric argument in favor of using weighted estimation for a common population coefficient if the underlying model varies per stratum (stratum-specific coefficient): both simple OLS and weighted OLS are inconsistent under regular conditions. However, weighted OLS has the advantage of providing an estimate that is independent of the sample design. This latter point may be quite relevant for the Enterprise Surveys as in most cases the objective is not only to obtain model-unbiased estimates but also design-unbiased estimates (see also Cochran, 1977, pp 200 who favors the used of weighted OLS for a common population coefficient.)<sup>6</sup>

29. From a more general approach, if the regressions are descriptive of the population then weights should be used. The estimated model can be thought of as the relationship that would be expected if the whole population were observed<sup>7</sup>. If the models are developed as structural relationships or behavioral models that may vary for different parts of the population, then, there is no reason to use weights.

## 8. Non-response

30. Survey non-response must be differentiated from item non-response. The former refers to refusals to participate in the survey altogether whereas the latter refers to the refusals to answer some specific questions. Enterprise Surveys suffer from both problems and different strategies were used to address these issues.

31. Item non-response was addressed by two strategies:

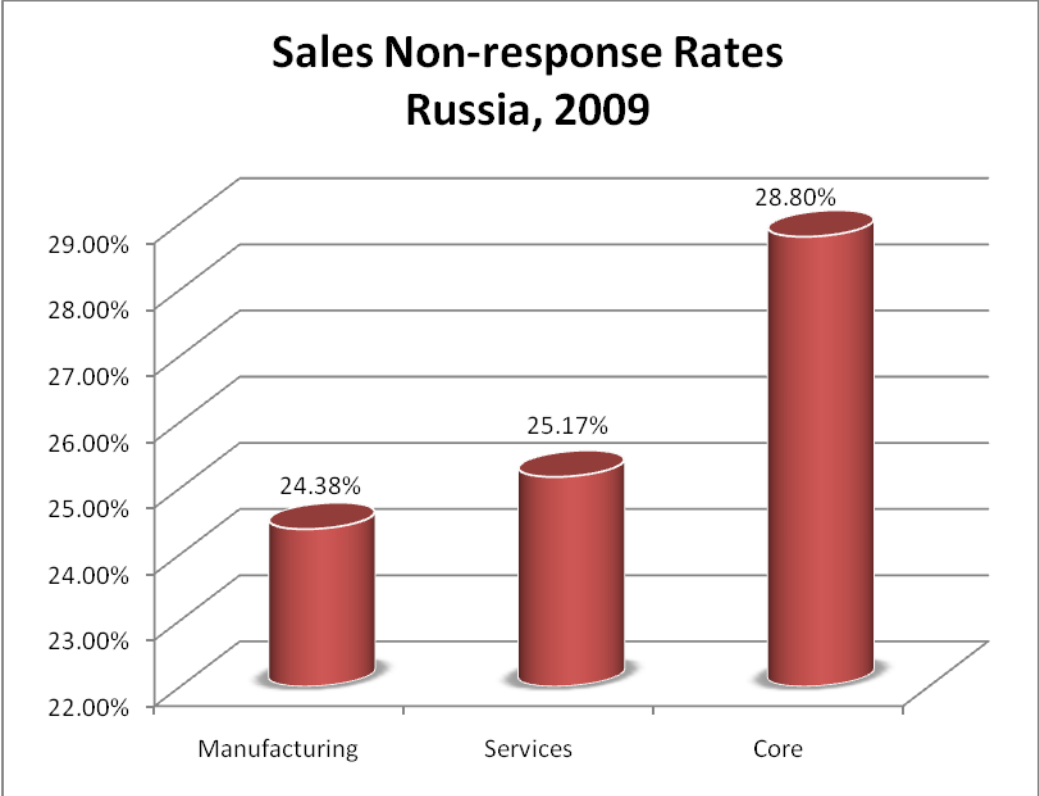
a- For sensitive questions that may generate negative reactions from the respondent, such as corruption or tax evasion, enumerators were instructed to collect the refusal to respond as (-8).

b- Establishments with incomplete information were re-contacted in order to complete this information, whenever necessary. However, there were clear cases of low response. The following graph shows non-response rates for the sales variable, *d2*, by type of questionnaire. Please, note that the coding utilized in this dataset does not allow us to differentiated between “Don’t know” and “refuse to answer”, thus the non-response in the table below reflects both categories (DKs and NAs).

---

<sup>6</sup> Note that weighted OLS in Stata using the command regress with the option of weights will estimate wrong standard errors. Using the Stata survey specific commands svy will provide appropriate standard errors.

<sup>7</sup> The use weights in most model-assisted estimations using survey data is strongly recommended by the statisticians specialized on survey methodology of the JPSM of the University of Michigan and the University of Maryland.



32. Survey non-response was addressed by maximizing efforts to contact establishments that were initially selected for interview. Up to 4 attempts were made to contact the establishment for interview at different times/days of the week before a replacement establishment (with similar strata characteristics) was suggested for interview. Survey non-response did occur but substitutions were made in order to potentially achieve strata-specific goals. Further research is needed on survey non-response in the Enterprise Surveys regarding potential introduction of bias.

33. As the following graph shows, the number of contacted establishments per realized interview was 6.14. This number is the result of two factors: explicit refusals to participate in the survey, as reflected by the rate of rejection (which includes rejections of the screener and the main survey) and the quality of the sample frame, as represented by the presence of ineligible units.



34. Details on rejections rates, eligibility rates, and item non-response are available at the level strata. This report summarizes these numbers to alert researchers of these issues when using the data and when making inferences. Item non-response, selection bias, and faulty sampling frames are not unique to Russia. All enterprise surveys suffer from these shortcomings but in very few cases they have been made explicit.

#### References

Cochran, William G., *Sampling Techniques*, 1977.

Deaton, Angus, *The Analysis of Household Surveys*, 1998.

Levy, Paul S. and Stanley Lemeshow, *Sampling of Populations: Methods and Applications*, 1999.

Lohr, Sharon L. *Sampling: Design and Techniques*, 1999.

Scheaffer, Richard L.; Mendenhall, W.; Lyman, R., *Elementary Survey Sampling*, Fifth Edition, 1996

## Appendix A Cell Weights – Russia (Strict)

Individual Cell Weights

Region	Employees	Sector								
		15	18	24	27-28	29	52	Other Manufacturing	Other Services	
North West	5 - 19	13	18	5	7	20	140	58	1,027	
	20 - 99	4	1	2	3	2	9	14	95	
	100+	2		1	3	3	18	19	535	
Central	5 - 19	6	1	3	12	10	108	55	489	
	20 - 99	2	1	2	2	2	18	22	171	
	100+	2	1	2	3	1	40	24	135	
South	5 - 19	9	2	5	9	17	133	125	631	
	20 - 99	4	1	3	2	3	19	44	153	
	100+	3	1		4	2	23		291	
Ural	5 - 19	20	2	7	12	14	111	90	482	
	20 - 99	4		1	3	5		28	408	
	100+	3		2	3	5		23	221	
Siberia	5 - 19	6	1	2	10	22	86	57	986	
	20 - 99	3		2	2	2	13	23	125	
	100+	3		2	2	7		12	151	
Volgo-Viatsky	5 - 19	9	2	8	20	25	225	110	1,043	
	20 - 99	4	2	2	4	4	29	38	258	
	100+	8		3	3	5	57	155	159	
Far East	5 - 19	2	2	2	3	2	25	13	147	
	20 - 99	2		1	2	2	20	4	93	
	100+	3				1	7		63	

Collapsed Cell Weights

Region	Employees	Sector								
		15	18	24	27-28	29	52	Other Manufacturing	Other Services	
North West	5 - 19	13	18	5	7	20	140	58	1,027	
	20 - 99	4	1	2	3	2	9	14	95	
	100+	2		1	3	3	18	19	535	
Central	5 - 19	6	1	3	12	10	108	55	489	
	20 - 99	2	1	2	2	2	20	22	171	
	100+	2	1	2	3	1	40	28	135	
South	5 - 19	9	2	5	9	17	133	125	631	
	20 - 99	4	1	3	2	3	19	44	153	
	100+	3	1		4	2	23		291	
Ural	5 - 19	20	2	7	12	14	111	90	482	
	20 - 99	4		1	3	5		28	408	
	100+	3		2	3	5		69	221	
Siberia	5 - 19	6	1	2	10	22	86	57	986	
	20 - 99	3		2	2	2	13	23	125	
	100+	3		2	2	7		12	151	
Volgo-Viatsky	5 - 19	9	2	8	20	25	225	110	1,043	
	20 - 99	4	2	2	4	4	29	38	258	
	100+	8		3	3	5	57	69	159	
Far East	5 - 19	2	2	2	3	2	25	13	147	
	20 - 99	2		1	2	2	20	4	93	
	100+	3				1	7		63	

## Cell Weights – Russia (Weak)

Individual Cell Weights

Region	Employees	Sector							
		15	18	24	27-28	29	52	Other Manufacturing	Other Services
North West	5 - 19	53	71	25	31	88	427	185	3,312
	20 - 99	15	2	8	13	9	26	43	288
	100+	8		3	10	11	47	54	1,489
Central	5 - 19	23	5	15	53	45	334	179	1,601
	20 - 99	8	5	10	9	8	52	69	522
	100+	6	3	8	10	5	107	68	380
South	5 - 19	20	4	11	22	40	219	218	1,097
	20 - 99	8	1	6	5	6	30	72	249
	100+	5	2		8	4	32		436
Ural	5 - 19	61	5	22	38	44	255	219	1,166
	20 - 99	12		2	10	14		63	923
	100+	9		7	7	15		48	462
Siberia	5 - 19	18	4	8	30	66	185	130	2,238
	20 - 99	9		5	4	7	27	48	265
	100+	7		5	4	18		23	295
Volgo-Viatsky	5 - 19	21	5	19	48	60	383	199	1,875
	20 - 99	8	3	4	8	8	46	64	433
	100+	15		6	6	10	84	240	247
Far East	5 - 19	3	4	3	5	4	30	16	185
	20 - 99	4		1	4	3	23	5	109
	100+	4				1	8		68

Collapsed Cell Weights

Region	Employees	Sector							
		15	18	24	27-28	29	52	Other Manufacturing	Other Services
North West	5 - 19	53	71	25	31	88	427	185	3,312
	20 - 99	15	2	8	13	9	26	43	288
	100+	8		3	10	11	47	54	1,489
Central	5 - 19	23	5	15	53	45	334	179	1,601
	20 - 99	8	5	10	9	8	57	69	522
	100+	6	3	8	10	5	107	72	380
South	5 - 19	20	4	11	22	40	219	218	1,097
	20 - 99	8	1	6	5	6	30	72	249
	100+	5	2		8	4	32		436
Ural	5 - 19	61	5	22	38	44	255	219	1,166
	20 - 99	12		2	10	14		63	923
	100+	9		7	7	15		113	462
Siberia	5 - 19	18	4	8	30	66	185	130	2,238
	20 - 99	9		5	4	7	27	48	265
	100+	7		5	4	18		23	295
Volgo-Viatsky	5 - 19	21	5	19	48	60	383	199	1,875
	20 - 99	8	3	4	8	8	46	64	433
	100+	15		6	6	10	84	113	247
Far East	5 - 19	3	4	3	5	4	30	16	185
	20 - 99	4		1	4	3	23	5	109
	100+	4				1	8		68

## Cell Weights – Russia (Median)

Individual Cell Weights

Region	Employees	Sector								
		15	18	24	27-28	29	52	Other Manufacturing	Other Services	
North West	5 - 19	25	33	11	14	42	223	102	1,699	
	20 - 99	8	1	4	7	5	16	27	171	
	100+	5		2	6	7	32	39	1,001	
Central	5 - 19	12	2	7	26	23	183	104	863	
	20 - 99	5	3	5	5	5	33	46	325	
	100+	4	2	5	7	3	76	51	268	
South	5 - 19	12	2	6	13	25	147	156	727	
	20 - 99	6	1	4	3	4	23	59	191	
	100+	4	2		7	3	29		378	
Ural	5 - 19	31	3	11	19	23	142	129	639	
	20 - 99	7		1	6	9		43	586	
	100+	6		4	5	10		38	332	
Siberia	5 - 19	11	2	4	18	40	124	92	1,475	
	20 - 99	6		3	3	5	21	39	202	
	100+	6		4	3	14		22	254	
Volgo-Viatsky	5 - 19	12	3	10	27	34	240	132	1,156	
	20 - 99	5	2	2	5	6	33	49	309	
	100+	11		4	4	8	69	209	199	
Far East	5 - 19	2	2	2	3	3	22	13	135	
	20 - 99	3		1	3	2	20	4	93	
	100+	4				1	7		66	

Collapsed Cell Weights

Region	Employees	Sector								
		15	18	24	27-28	29	52	Other Manufacturing	Other Services	
North West	5 - 19	25	33	11	14	42	223	102	1,699	
	20 - 99	8	1	4	7	5	16	27	171	
	100+	5		2	6	7	32	39	1,001	
Central	5 - 19	12	2	7	26	23	183	104	863	
	20 - 99	5	3	5	5	5	36	46	325	
	100+	4	2	5	7	3	76	56	268	
South	5 - 19	12	2	6	13	25	147	156	727	
	20 - 99	6	1	4	3	4	23	59	191	
	100+	4	2		7	3	29		378	
Ural	5 - 19	31	3	11	19	23	142	129	639	
	20 - 99	7		1	6	9		43	586	
	100+	6		4	5	10		96	332	
Siberia	5 - 19	11	2	4	18	40	124	92	1,475	
	20 - 99	6		3	3	5	21	39	202	
	100+	6		4	3	14		22	254	
Volgo-Viatsky	5 - 19	12	3	10	27	34	240	132	1,156	
	20 - 99	5	2	2	5	6	33	49	309	
	100+	11		4	4	8	69	96	199	
Far East	5 - 19	2	2	2	3	3	22	13	135	
	20 - 99	3		1	3	2	20	4	93	
	100+	4				1	7		66	

## Appendix B

### Status Codes - Total

ELIGIBLES		
Eligible	1. Eligible establishment ( <i>Correct name and address</i> )	1784
	2. Eligible establishment ( <i>Different name but same address - the new firm/establishment bought the original firm/establishment</i> )	3
	3. Eligible establishment ( <i>Different name but same address - the firm/establishment changed its name</i> )	8
	4. Eligible establishment ( <i>Wrong address - the firm/establishment has changed address and the address could be found</i> )	47
	16. Panel firm - now less than five employees	4
Ineligible	5. The establishment has less than 5 permanent full time employees	35
	6. The firm discontinued businesses	125
	7. Not a business: private household	28
	8. Ineligible activity: education, agriculture, finances, governments...	85
Unobtainable	91. No reply ( <i>after having called in different days of the week and in different business hours</i> )	833
	92. Line out of order	188
	93. No tone	20
	10. Answering machine	13
	11. Fax line - data line	34
	12. Wrong address/ moved away and could not get the new references	1097
	13. Refuses to answer the screener	1245
	<b>14. In process</b> ( <i>the establishment is being called/ is being contacted - previous to ask the screener</i> )	<b>625</b>
	151. Out of target - outside the covered regions, firm moved abroad	10
	152. Out of target - firm moved abroad	0
<b>Total</b>		<b>6184</b>

### Response Outcomes - Total

Complete interviews ( <i>Total</i> )	1004
Incomplete interviews	24
Eligible in process	106
Refusals	712
Out of target	273
Impossible to contact	2185
Ineligible - coop.	10
Refusal to the Screener	1245
<b>Total</b>	<b>9559</b>

### Status Codes – Panel

Eligibles		
Eligible	1. Eligible establishment (Correct name and address)	77
	2. Eligible establishment (Different name but same address - the new firm/establishment bought the original firm/establishment)	1
	3. Eligible establishment (Different name but same address - the firm/establishment changed its name)	3
	4. Eligible establishment (Wrong address - the firm/establishment has changed address and the address could be found)	10
	16. Panel firm - now less than five employees	2
Ineligible	5. The establishment has less than 5 permanent full time employees	1
	6. The firm discontinued businesses	12
	7. Not a business: private household	0
	8. Ineligible activity: education, agriculture, finances, governments...	6
Unobtainable	91. No reply ( <i>after having called in different days of the week and in different business hours</i> )	47
	92. Line out of order	24
	93. No tone	1
	10. Answering machine	0
	11. Fax line - data line	4
	12. Wrong address/ moved away and could not get the new references	59
	13. Refuses to answer the screener	30
	<b>14. In process (<i>the establishment is being called/ is being contacted - previous to ask the screener</i>)</b>	<b>46</b>
	151. Out of target - outside the covered regions	8
	152. Out of target - firm moved abroad	0
	<b>Total</b>	<b>331</b>

### Response Outcomes – Panel

Complete interviews (Total)	57
Incomplete interviews	4
Elegible in process	6
Refusals	26
Out of target	19
Impossible to contact	135
Ineligible - coop.	8
Refusal to the Screener	30
<b>Total</b>	<b>285</b>



## Status Codes – ROMIR Fresh

Eligibles		
Eligible	1. Eligible establishment (Correct name and address)	130
	2. Eligible establishment (Different name but same address - the new firm/establishment bought the original firm/establishment)	0
	3. Eligible establishment (Different name but same address - the firm/establishment changed its name)	0
	4. Eligible establishment (Wrong address - the firm/establishment has changed address and the address could be found)	3
	16. Panel firm - now less than five employees	0
Ineligible	5. The establishment has less than 5 permanent full time employees	0
	6. The firm discontinued businesses	10
	7. Not a business: private household	16
	8. Ineligible activity: education, agriculture, finances, governments...	3
Unobtainable	91. No reply ( <i>after having called in different days of the week and in different business hours</i> )	113
	92. Line out of order	10
	93. No tone	1
	10. Answering machine	1
	11. Fax line - data line	4
	12. Wrong address/ moved away and could not get the new references	101
	13. Refuses to answer the screener	86
	<b>14. In process (<i>the establishment is being called/ is being contacted - previous to ask the screener</i>)</b>	<b>128</b>
	151. Out of target - outside the covered regions	0
	152. Out of target - firm moved abroad	0
	<b>Total</b>	<b>606</b>

## Response Outcomes – ROMIR Fresh

Complete interviews (Total)	83
Incomplete interviews	9
Elegible in process	32
Refusals	9
Out of target	29
Impossible to contact	230
Ineligible - coop.	0
Refusal to the Screener	86
<b>Total</b>	<b>478</b>

### Status Codes – Fresh

Eligibles		
Eligible	1. Eligible establishment (Correct name and address)	1577
	2. Eligible establishment (Different name but same address - the new firm/establishment bought the original firm/establishment)	2
	3. Eligible establishment (Different name but same address - the firm/establishment changed its name)	5
	4. Eligible establishment (Wrong address - the firm/establishment has changed address and the address could be found)	34
	16. Panel firm - now less than five employees	2
Ineligible	5. The establishment has less than 5 permanent full time employees	34
	6. The firm discontinued businesses	103
	7. Not a business: private household	12
	8. Ineligible activity: education, agriculture, finances, governments...	76
Unobtainable	91. No reply ( <i>after having called in different days of the week and in different business hours</i> )	673
	92. Line out of order	154
	93. No tone	18
	10. Answering machine	12
	11. Fax line - data line	26
	12. Wrong address/ moved away and could not get the new references	937
	13. Refuses to answer the screener	1129
	<b>14. In process (<i>the establishment is being called/ is being contacted - previous to ask the screener</i>)</b>	<b>451</b>
	151. Out of target - outside the covered regions	2
	152. Out of target - firm moved abroad	0
	<b>Total</b>	<b>5247</b>

### Response Outcomes – Fresh

Complete interviews (Total)	866
Incomplete interviews	9
Elegible in process	68
Refusals	677
Out of target	225
Impossible to contact	1820
Ineligible - coop.	2
Refusal to the Screener	1129
<b>Total</b>	<b>4796</b>

## Appendix C

### Eligibility Rules

Status Code	Eligibility Criteria		
	Strict	Weak	Median
1. Eligible establishment (Correct name and address)	1	1	1
2. Eligible establishment (Different name but same address - the new firm/establishment bought the original firm/establishment)	1	1	1
3. Eligible establishment (Different name but same address - the firm/establishment changed its name)	1	1	1
4. Eligible establishment (Wrong address - the firm/establishment has changed address and the address could be found)	1	1	1
16. Panel firm - now less than five employees	1	1	1
5. The establishment has less than 5 permanent full time employees	0	0	0
6. The firm discontinued businesses	0	0	0
7. Not a business: Private household	0	0	0
8. Ineligible activity: education, agriculture, finances, governments...	0	0	0
91. No reply (after having called in different days of the week and in different business hours)	0	1	0
92. Line out of order	0	1	0
93. No tone	0	1	0
10. Answering machine	0	1	1
11. Fax line – data line	0	1	1
12. Wrong address/ moved away and could not get the new references	0	1	0
13. Refuses to answer the screener	0	1	1
<b>14. In process (the establishment is being called/ is being contacted – previous to ask the screener)</b>	0	0	0
151. Out of target – outside the covered regions, firm moved abroad	0	0	0
152. Out of target – firm moved abroad	0	0	0

Strict eligibility

= (Sum of the numbers with codes 1,2,3,4,&16) / Total

Weak eligibility

= (Sum of the numbers with codes 1,2,3,4,16,91,92,93,10,11,12,&13) / Total

Median eligibility

= (Sum of the numbers with codes 1,2,3,4,16,10,11, & 13) / Total

### Russia Establishment Estimates

Cells	Strict	Weak	Median
Un-collapsed Cells	56,968	146,942	87,806
Collapsed Cells	57,069	147,074	87,925

## Appendix D

### Questionnaires:

Problems for the understanding of questions (write question number)	<p>Sometimes there was confusion surrounding the questions on unofficial gifts. No additional explanation was given by the interviewer given the sensitivity of such questions and the question was read exactly as it appeared on the questionnaire.</p> <p>Talking about water supply costs, costs for electricity etc. was problematic in numerous cases. In cases where the establishment does not have its own premises but rents a premises or office respondents could not indentify the individual costs for all statements as the costs for renting includes costs for water supply, electricity and others.</p>
Problems found in the navigability of – questionnaires (for example, skip patterns).	No problems. Perhaps questions about financial statements should be asked earlier (in the middle of interview). Sometimes people have no patience to complete the full interview and moreover share the company's financial information.
Comments on questionnaires length:	The interview is lengthy and sometimes it was difficult to reach the end of the interview.
Suggestions or other comments on the questionnaire:	As the questionnaire length is very long self-completion could perhaps be considered in the future.

### Database

Comments on the data entry program	<p>Data entry program chosen: Confirmit (TNS MIC) &amp; CENTRY (ROMIR)</p> <p>Comments on Confirmit: On the one hand the fact that it's impossible to skip answers is very useful for data entry department as they can't miss any answer. On the other hand, the local institute faced the challenge of quantitative questions, sometimes respondent gave an answer that contains strange numbers in terms of Confirmit (e.g. amount of certificates, licence, etc.). Additional data checking instructions were implemented. The advantage of the programme was that it could identify the mistakes at the very moment the questionnaire was entered or show the list of errors after it was entered. In that case we could make the call-backs and corrections immediately and not wait for the data validation report.</p> <p>Comments on CENTRY: None</p>
Comments on the data cleaning	<p>Comments from TNS MIC: Working with the data validation reports was a difficult and time consuming process.</p> <p>Comments from ROMIR: None</p>

### Country situation

General aspects of economic, political or social situation of the country that could affect the results of the survey:	During the fieldwork and even prior to commencing, many Russian factories were closed because of the global financial crisis.
Relevant country events occurred during fieldwork:	A lot of employees were fired in many companies, so the information about the number of employees was rapidly evolving.
Other aspects:	None

**Appendix E  
Original Sample Design**

Region	Employees	Sector										Grand Total
		15	18	24	27-28	29	52	Manufacturing	Other Services			
North West	5 - 19	8	15	7	11	9	6	7	5	68		
	20 - 99	10	7	11	12	10	4	6	4	64		
	100+	9	2	7	7	8	5	6	6	50		
North West Total		27	24	25	30	27	15	19	15	182		
Central	5 - 19	27	67	30	22	26	20	21	23	236		
	20 - 99	21	16	29	21	24	22	22	23	178		
	100+	24	8	22	17	14	8	25	20	138		
Central Total		72	91	81	60	64	50	68	66	552		
South	5 - 19	5	8	2	4	3	5	2	2	31		
	20 - 99	6	1	1	3	4	6	2	3	25		
	100+	3	1	1	3	3	2	1	2	16		
South Total		14	9	4	10	10	13	5	7	72		
Ural	5 - 19	3	7	3	5	5	3	2	3	31		
	20 - 99	4	2	1	5	4	2	2	2	22		
	100+	3	3	3	11	9	5	2	2	35		
Ural Total		10	9	7	21	18	10	6	7	88		
Siberia	5 - 19	5	7	3	4	3	5	3	2	32		
	20 - 99	4	3	3	4	3	7	2	2	25		
	100+	4	2	2	6	5	1	1	3	21		
Siberia Total		13	7	8	14	11	12	6	7	78		
Volgo-Viatsky	5 - 19	7	17	10	9	9	9	5	5	71		
	20 - 99	10	3	10	10	10	6	6	6	61		
	100+	7	15	15	6	11	5	5	7	56		
Volgo-Viatsky Total		24	20	35	25	30	20	16	18	188		
Far East	5 - 19	12	4	2	3	6	20	16	10	73		
	20 - 99	3	1	1	1	1	2	5	4	16		
	100+	4	4	2	4	7	2	2	5	11		
Far East Total		19	4	2	4	7	24	21	19	100		
Grand Total		179	164	162	164	167	144	141	139	1260		

## Appendix F

### Local Agency teams involved in the study:

Local Agency	Name: TNS MIC Country: Russia Membership of international organisation: ESOMAR Activities since: 1990 (part of TNS since 2001)
Name of Project Manager	Boris Khatutsky
Name and position of other key persons of the project:	Dmitriy Trofimov, Head of Research Group Tatyana Apraksina, Manager
Enumerators involved:	Enumerators: 88 Recruiters: 77 49 people were both enumerators and recruiters.
Other staff involved:	Fieldwork Coordinators: Elena Sulkova Field Manager: Igor Zakurdaev Editing: Maria Pirojkova Data Entry: Tatyana Sysoeva Data Processing:

Local Agency	Name: ROMIR Country: Russia Membership of international organisation: Gallup International, Global NR, Open World and Worldwide Independent Network (WIN) Activities since: 1987
Name of Project Manager	Belackovskaya Natalia
Name and position of other key persons of the project:	Shakaryan Aida – customer service manager
Enumerators involved:	Enumerators: 67 interviewers and 9 supervisors Recruiters: Interviewers also did recruitment.
Other staff involved:	Fieldwork Coordinators: Adonieva Madina Editing: Data Entry: Elena Lyalek Data Processing: Roman Miniahmetov

### Sample Frame:

Characteristic of sample frame used:	GosKomStat is an official source of information. The information is based on the financial statements of the companies (book-keeping reports).
Source:	GosKomStat - National Statistics Agency
Year of publication:	Beginning of 2007
Comments on the quality of sample frame:	Many incorrect telephone numbers and addresses for the reason that most companies have different official address than the actual location of the establishments' activities. GosKomStat's base includes official addresses only. The local institutes had to check every second contact and much time was dedicated to searching for updated contact information. However, this is the unique source of information available in Russia in order to have a data base which is representative of the target universe.
Year and organism who conducted the last economic census	It is updated yearly. The version used was updated at the beginning of 2007 so the contact information refers to 2006.
Other sources for companies statistics	Online search for actual addresses and telephone numbers (including the website: <a href="http://yellowpages.ru">http://yellowpages.ru</a> ), several subcontractors used their own data bases to find appropriate information about firms from the sample (in some regions). This approach was used in all the cases when we could reach companies using the contact information from the initial sample frame.

### Sample:

Comments/ problems on sectors and regions selected in the sample:	On sectors: Industry classification used in the sample frame has a different name (OKVED) however it is the same as ISIC. On regions: Central region was the most difficult one because of the large target sample. The region is mostly represented by Moscow city only so many interviews had to be done in Moscow.
Comments on the response rate:	Only 14,000 contacts were eligible for purchase from the GosKomStat sample frame for the 53 cities targeted. The poor quality of the sample frame in terms of the contact information contributed to a low response rate.
Comments on the sample design:	Size of establishment was estimated from the turnover as it was not indicated in the initial sample frame.
Other Comments:	All establishments which have their own financial statements are included given that the criterion to be included in the GosKomStat base is to report financial statements.



## Appendix G. Fieldwork:

Date of Fieldwork	5 <sup>th</sup> of September 2008 - 24 <sup>th</sup> of March 2009
Country	Russia
Interview number	Manufactures: 630 Services: 111 Core: 263
Problems found during fieldwork:	<p>The low quality of the sample frame (too many companies which do not exist or contact information is inaccurate).</p> <p>Respondents did not see how participation in the study will be advantageous for them; they do not believe that the World Bank's policy will help them directly.</p> <p>The interview is very lengthy.</p> <p>Giving financial information during the interview was a real barrier for cooperation in spite of interviewers' efforts to reassure respondents that their responses would never be related back to them individually.</p> <p>Respondents were fearful nonetheless and considered that it was their 'company secret'. Sometimes it was a reason for refusal at the very beginning of the interview.</p>
Other observations:	<p>The financial crisis in Russia influenced the response rate. All the companies' top-managers were preoccupied with company business and had neither the time nor the wish to participate. Generally-speaking they are not confident in the future. Given their focus on resolving problems which the financial crisis brings to their company, they were unwilling to dedicate time to an interview.</p> <p>There were some cases when respondents were willing to complete the questionnaire themselves and send it by e-mail. Many respondents were willing to answer all the questions during a telephone call but not to meet the interviewer.</p>

## **Appendix H. Survey Universe, Sample Population and Sampling Frames**

The following provides description of the general methodology used in BEEPS 2009.

The survey universe was defined as commercial, service or industrial business establishments with at least five full-time employees. Government departments including military, police, education, health and similar activities were excluded, as were those in primary industries including agriculture, mining, etc.

There are no up to date and reliable statistics relating to this universe in the countries being surveyed in BEEPS IV. Consequently the universe size and characteristics have to be directly estimated from the survey results themselves. This requirement increases the emphasis that has to be placed on the quality of the sample frame, because the validity of the results is predominantly a function of coverage and age of the sampling frame.

The criteria used to evaluate the available sampling frame in descending priority were those of:

- Coverage

- ✓ Up to datedness
- ✓ Availability of detailed stratification variables
- ✓ Location identifiers- address, phone number, email
- ✓ Electronic format availability
- ✓ Contact name(s)

The sample frames used for the surveys must consist of the lists of enterprises in each country that most optimally meet these requirements. The final selection was made by the TNS in collaboration with the World Bank and EBRD. For most countries covered in BEEPS IV two sample frames were used. The first frame was often an official frame of establishments supplied by the national statistical office of the country. The Enterprise Survey conducted for the World Bank in Albania in 2007/8 showed that a suitable frame did not exist for the country. Instead, the design returned to first principles, using a blocks enumeration methodology.